The application of particle filters in single trial event-related potential estimation

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

2009 Physiol. Meas. 30 1101

(http://iopscience.iop.org/0967-3334/30/10/010)

View the table of contents for this issue, or go to the journal homepage for more

Download details:
IP Address: 163.1.42.75
The article was downloaded on 19/04/2010 at 09:53

Please note that terms and conditions apply.
The application of particle filters in single trial event-related potential estimation

Hamid R Mohseni1,2, Kianoush Nazarpour1,3, Edward L Wilding2 and Saeid Sanei1

1 School of Engineering, Cardiff University, Cardiff, UK
2 Cardiff University Brain Research Imaging Centre, School of Psychology, Cardiff University, Cardiff, UK
3 School of Psychology, University of Birmingham, Birmingham, UK

E-mail: K.Nazarpour@bham.ac.uk

Received 30 March 2009, accepted for publication 7 August 2009
Published 16 September 2009
Online at stacks.iop.org/PM/30/1101

Abstract
In this paper, an approach for the estimation of single trial event-related potentials (ST-ERPs) using particle filters (PFs) is presented. The method is based on recursive Bayesian mean square estimation of ERP wavelet coefficients using their previous estimates as prior information. To enable a performance evaluation of the approach in the Gaussian and non-Gaussian distributed noise conditions, we added Gaussian white noise (GWN) and real electroencephalogram (EEG) signals recorded during rest to the simulated ERPs. The results were compared to that of the Kalman filtering (KF) approach demonstrating the robustness of the PF over the KF to the added GWN noise. The proposed method also outperforms the KF when the assumption about the Gaussianity of the noise is violated. We also applied this technique to real EEG potentials recorded in an odd-ball paradigm and investigated the correlation between the amplitude and the latency of the estimated ERP components. Unlike the KF method, for the PF there was a statistically significant negative correlation between amplitude and latency of the estimated ERPs, matching previous neurophysiological findings4.

Keywords: event-related potentials, Kalman filtering, particle filtering, single trial estimation

(Some figures in this article are in colour only in the electronic version)

---

1. Introduction

Event-related potentials (ERPs) are the most widely employed real-time measure of neural activity (Rugg and Coles 1995). Conventional methods for analysing ERPs typically involve time-locked averaging over many trials. The assumption underlying this approach is that the background EEG—a random process—will be attenuated by averaging. If the results of averaging are to be an accurate reflection of the activity elicited on individual trials, then positive- and negative-going ERP modulations (components) on each trial must have the same onset latencies, durations and amplitudes. Recent studies (Jongsma et al 2006) have indicated that there is single-trial variability in ERPs due to environmental and cognitive factors that might include fatigue, habituation and changes in levels of attention. These factors are of course not mutually exclusive, but the key point here is that the referenced work emphasizes that single-trial ERP (ST-ERP) mining offers to capture neurophysiological signal changes of interest that are lost when using conventionally averaged measures. Our primary objective here is to develop a reliable ST-ERP estimation method to detect and interpret ERP amplitude and latency variability during the course of a recording session.

Pioneering studies such as (McGillem and Aunon 1977) employed a time-invariant linear minimum mean squared error filter to characterize the ST-ERPs based on the mean, auto- and cross-covariance values of the background EEG (as noise) and ERPs (as signals of interest). Statistical signal processing methods including maximum a posteriori (MAP) solutions (Truccolo et al 2003) and Wiener filtering (Cerutti et al 1983) are the other approaches that have been widely utilized. In Pham et al (1987) and Jaskowski and Verleger (1999) the maximum likelihood (ML) approach was formulated, yielding estimators for amplitude and latency variability (jitter). The pitfall of such approaches, however, lies in considering the ERP signal as a stationary process: ERPs are superpositions of transient responses with changing temporal and spectral components.

An extension of the ML model was applied in de Munck et al (2004) which estimated the ERPs from multichannel EEG where the ongoing brain activity was modelled with a stationary autoregressive (AR) time series. Such an AR modelling was also utilized in Xu et al (2009). A Bayesian approach for realizing a robust ST estimator for ERPs was proposed in Karjalainen et al (1999). In this approach, using subspace regularization, the second-order statistical information extracted from a set of measured potentials was used to represent a priori knowledge in the estimation procedure. It was shown that the latencies of the ERP peaks can be estimated accurately. This approach does not, however, yield reliable estimates of peak amplitudes, mainly due to unpredictable fluctuations of the background EEG. A related work was carried out in Ranta-aho et al (2003).

Kalman filtering (KF) has also been developed for the separation of each single measured response (trial) into background activity and ERP components (Spreckelsen and Bromm 1988). Background EEG was modelled as an AR process and estimated from the activity measured immediately before each stimulus, while the ERPs were considered as the output of a linear system with an impulse signal as the input. KF has also been employed for estimation of the amplitude and latency changes of ERPs from trial to trial by considering the entire trial of the ERP as the input (Georgiadis et al 2005). In this case, however, high dimensionality of the input would impair the performance of the filter. Likewise, Kalman smoothing has been applied to ST-ERP, and the results were compared with those of the conventional averaging method (Mohseni et al 2007).

Independent component analysis (ICA) based methods have also been widely exploited in ERP extraction (Jung et al 2001, Lemm et al 2006, Sanei and Chambers 2007). This approach assumes that the EEG sources are independent, which can be violated in the case of low
signal-to-noise ratio ERPs and also when correlated noise sources are present. Other
techniques that have been employed include parametric methods (Lange and Inbar 1996,
Lange et al 1997), matched filters (Lange et al 1995) and discrete wavelet transforms (DWT)
(Bartnik et al 1999, Quiroga and Garcia 2003). The latter is able to accommodate the
non-stationarity of ERPs (see section 2).

In this study, we present an algorithm for estimation of ST-ERPs using particle filters
(PFs). The PF is an emerging and powerful technique for sequential signal processing with a
wide range of applications in bioinformatics, medicine and engineering. In our experiments,
as will be discussed later, if a sufficient number of particles are chosen, PF outperforms the
KF. In addition, PF is a sequential method and can track variability of trials (as can KF),
which represents another advantage over some other methods, in particular ML, ICA and
DWT. However, in comparison with other methods, such as ICA, PF does not use the spatial
information. Also in contrast to MAP estimation, PF does not use any prior knowledge about
the measurements.

This paper is organized as follows. First, we formulate the ST-ERP estimation in the state
space. Kalman- and particle filter-based recursive solutions to this problem are the subject of
sections 2.1 and 2.2. In Section 3, we present the results and compare the performance of the
KF and PF methods. Section 4 contains concluding remarks.

2. Problem formulation in state space

The performance of PF and KF estimators in the state space is highly influenced by the
dimension of the hidden state vector \( \mathbf{x} \). In order to reduce this dimension, one could
use the DWT, which decomposes the signal into high (detail) and low (approximation)
frequency contained coefficients by incorporating a set of predefined wavelets (Mallat 1999).
Considering that ERPs have relatively low-frequency spectral components with different
temporal localizations, DWT is recommended for denoising and dimension reduction in
(Jansen 2001). Therefore, let \( m \) approximation wavelet coefficients of time locked measured
ERPs in the \( k \)th (\( k \in \mathbb{N} \)) trial be as
\[
\mathbf{y}_k = [y_k(1) y_k(2) \cdots y_k(m)]^T,
\]
where \([ \cdot ]^T\) denotes the transpose operation. By modelling the wavelet coefficients in the state
space, the evolution of the state \( \mathbf{x}_k \) and the relations between the states and the measurements
(wavelet coefficients) are, respectively, given by
\[
\mathbf{x}_k = f_{k-1}(\mathbf{x}_{k-1}, \mathbf{w}_{k-1})
\]
\[
\mathbf{y}_k = h_k(\mathbf{x}_k, \mathbf{v}_k)
\]
where \( f_k \) and \( h_k \) are functions applied to the hidden state \( \mathbf{x}_k \), and \( \mathbf{w}_{k-1} \) and \( \mathbf{v}_k \) are assumed
to be zero mean Gaussian white noise (GWN) with known covariance matrices \( Q_w \) and \( Q_v \),
respectively. A widely used model for ERP estimation is the linear additive noise model. The
measurement equation (3) is then assumed to be of the form
\[
\mathbf{y}_k = \mathbf{x}_k + \mathbf{v}_k,
\]
where the state \( \mathbf{x}_k \) corresponds to the approximation of the wavelet coefficients of the activity
that is related to the measured ERPs. The noise \( \mathbf{v}_k \) can be either assumed as GWN or the
background EEG independently distributed from the ERPs. We search for the filtered estimates
of \( \mathbf{x}_k \) based on a set of available measurements \( \mathbf{y}_{1:k} = \{y_i, i = 1, \ldots, k\} \), up to the \( k \)th trial.
Thus, one can estimate the posteriori density function \( p(\mathbf{x}_k|\mathbf{y}_{1:k}) \) of state \( \mathbf{x}_k \), and the estimation
of state $x_k$ will be the expected value of the posteriori density. Via the Bayes rule, the available measurement $y_k$ is used to update the posteriori density as

$$p(x_k|y_{1:k}) = \frac{p(y_k|x_k)p(x_k|x_{k-1})}{p(y_k|y_{1:k-1})}p(x_k|y_{1:k-1}).$$

(5)

We therefore sequentially estimate the approximate wavelet coefficients of ERPs which have been smoothed by zeroing the detail coefficients. ERPs are then reconstructed by taking the inverse wavelet transforms of the estimated approximate wavelet coefficients.

Since the transition from $p(x_k-1|y_{1:k-1})$ to $p(x_k|y_{1:k})$ is often analytically intractable, Kalman and particle filtering and their variations are the two major approaches to solve equation (5) recursively. In the following sections, these algorithms are reviewed and compared.

2.1. Kalman filter

The KF is an efficient recursive filter that estimates the state of a dynamic system from a series of noisy measurements. In KF, the evolution of the states and the relation between the measurements and states are considered to be linear. Equations (2) and (3) are therefore rewritten as

$$x_k = F_{k-1}x_{k-1} + w_{k-1}$$

(6)

$$y_k = H_kx_k + v_k.$$  

(7)

In addition, the posteriori distribution (state distribution) is considered to be Gaussian, so it can be characterized by its mean vector and covariance matrix. Thus, only the mean vector and the covariance matrix are calculated in each step, and one can evaluate them in a closed form as in Haykin (1996),

$$m_{k,k-1} = F_{k-1}m_{k-1,k-1}$$

(8)

$$P_{k,k-1} = Q_w + F_{k-1}P_{k-1,k-1}F_{k-1}^T$$

(9)

$$m_{k,k} = m_{k,k-1} + K_k(y_k - H_km_{k,k-1})$$

(10)

$$P_{k,k} = P_{k,k-1} - K_kA_{k-1}K_k^T,$$  

(11)

where $m_{i,j}$ and $P_{i,j}$ are, respectively, the mean vector and covariance matrices of the conditional probability density function (pdf) of $p(x_i|y_j)$, $A_{k-1} = H_kP_{k,k-1}H_k^T + Q_v$ is the covariance matrix of the innovation term $y_k - H_km_{k,k-1}$, and $K_k = P_{k,k-1}H_k^T H_k + S_{k-1}$ is the Kalman gain.

2.2. Particle filters

In a PF, the posteriori distributions are approximated by discrete random measures defined by particles $\{x^{(n)}, n = 1, \ldots, N\}$ and their associated weights $\{w^{(n)}, n = 1, \ldots, N\}$, where $N$ is the number of particles. The distribution based on these samples and weights is approximated as

$$p(x) \approx \sum_{n=1}^N w^{(n)}\delta(x - x^{(n)}),$$

(12)

where $\delta$ is the Dirac delta function. By this formulation, the posteriori distribution can have any form and is not necessarily Gaussian.
When generation of the particles by direct sampling from an unknown distribution \( p(x) \) is impossible, the particles are generated from a known distribution \( \pi(x) \) called importance density. This concept is known as importance sampling and results in the following estimates of the weights (Arulampalam et al 2002):

\[
w^{(n)}_k \propto \frac{p(x^{(n)}_k)}{\pi(x^{(n)}_k)}. \tag{13}
\]

Suppose at the \( k \)-th trial, the posteriori distribution \( p(x_k|y_{1:k}) \) is to be estimated subject to \( p(x_{k-1}|y_{1:k-1}) \). Given the discrete random measure \( \{x^{(n)}_{k-1}, w^{(n)}_{k-1}; n = 1, \ldots, N \} \) and the observation \( y_k \), one should approximate \( \{x^{(n)}_k, w^{(n)}_k; n = 1, \ldots, N \} \). If the importance density is chosen such that it is factorized to the distribution, the better the approximation will be. The choice of importance density is crucial in designing a PF. Notably, the closer the importance of the particles. We have followed the approach proposed in Liu (1996).

\[
x^{(n)}_k \propto \frac{p(x^{(n)}_k|y_{1:k}) p(x^{(n)}_k)}{\pi(x^{(n)}_k|y_{1:k})} \tag{15}
\]

The choice of importance density is crucial in designing a PF. Notably, the closer the importance density to the distribution, the better the approximation will be. The most popular choice for the importance density, which was also exploited in this study, is \( \pi(x_k|x^{(n)}_{k-1}, y_{1:k}) = p(x_k|x^{(n)}_{k-1}) \). This implies that equation (15) simply reduces to

\[
w^{(n)}_k \propto w^{(n)}_{k-1} p(y_k|x^{(n)}_k). \tag{16}
\]

This choice of importance density needs to sample from \( p(x_k|x^{(n)}_{k-1}) \), and therefore a sample can be obtained by generating the noise sample \( w^{(n)}_{k-1} \sim N(0, Q_w) \), where \( N(0, Q_w) \) is a zero mean Gaussian distribution with covariance matrix \( Q_w \), and setting \( x^{(n)}_k = f_{k-1}(x^{(n)}_{k-1}, w^{(n)}_{k-1}) \). Such an importance density function allows the weights to be readily computed as samples from this density. On the other hand, as this choice of importance density is independent from the measurements, the state space is explored without any knowledge of the observations. Therefore, this filter can suffer from the presence of outliers.

**Resampling:** The importance sampling weights indicate the importance level of the corresponding particles. A relatively small weight implies that the particle is drawn far from the main body of the posterior distribution and it makes only a small contribution to the final estimation. An increment in the number of particles with small weights degenerates the filtering performance. This can be compensated for by eliminating those particles with small weights and replicating ones with larger weights. This procedure is termed resampling. Resampling terminates when all weights are equal to \( 1/N \). The number of particles denotes the importance of the particles. We have followed the approach proposed in Liu (1996).

In algorithm 1, the pseudo-code of the PF for ST-ERP estimation is presented. Note that, because the resampling is involved in each step of filtering, equation (16) is reduced to

\[
w^{(n)}_k = p(y_k|x^{(n)}_k). \]
updated readily as \( w^{(n)}_k = N(\mathbf{y}_k | \mathbf{x}^{(n)}_k, Q_v) \), where \( N(\mathbf{a} | \mu, Q) \) denotes the Gaussian density function with mean vector \( \mu \) and covariance matrix \( Q \), evaluated at \( \mathbf{a} \).

Algorithm 1 Pseudocode for particle filtering
Set \( k = 0 \) and generate random numbers \( \mathbf{x}_0^{(n)} \) according to a random uniform distribution
\[ \text{for } k = 1 \text{ to } \mathcal{T} \text{ do } \{ \mathcal{T} \text{ is the number of trials} \} \]
- Generate random numbers \( \mathbf{w}^{(n)}_{k-1} \sim N(0, Q_w) \) and set \( \tilde{\mathbf{x}}^{(n)}_k = f(\mathbf{x}^{(n)}_{k-1}, \mathbf{w}^{(n)}_{k-1}) \).
- Update new weights \( w^{(n)}_k = N(\mathbf{y}_k | \tilde{\mathbf{x}}^{(n)}_k, Q_v) \).
- Normalize the weights \( \sum_{n=1}^{\mathcal{N}} w^{(n)}_k = 1 \).
- Resample new \( \mathcal{N} \) particles \( \tilde{\mathbf{x}}^{(n)}_k \) from \( \tilde{\mathbf{x}}^{(n)}_k \) with replacement according to the importance weights \( w^{(n)}_k \) (see the text).
- Set \( w^{(n)}_k = \frac{1}{\mathcal{N}} \).
\[ \text{end for} \]

3. Materials and results
In this section, comparisons between PF and KF for simulated ERPs are given first, and then we report the results of applying both algorithms to real EEG measurements obtained in an odd-ball paradigm.

3.1. Simulation results
To generate a set of synthetic EEG data, we followed the approach introduced in Georgiadis et al. (2005). Simulated EEGs contained ERP waves in the interval between 0.2 s and 0.5 s post-stimulus. The sampling frequency and the number of trials were, respectively, set to 250 Hz and 60. These values were chosen in order to match those for the real EEG data set to which the approach was applied subsequently. Two Gaussian functions (one positive and one negative) were used to simulate the ERPs. The amplitude, mean and variance of Gaussian functions represent the amplitude, latency and width of the ERP components, respectively. The amplitude and latency of the positive peak were modelled to respectively decrease and increase linearly during the course of the recording from trial to trial. GWN with different levels was added to both the amplitudes and the latencies. The amplitude and latency of the negative peak were assumed to vary according to a uniform random distribution. The widths of both positive and negative peaks were constant from trial to trial; however, some normally distributed random variations were added to them.

For better assessment of the proposed algorithms, the GWN and real background EEG activities were considered as two different types of random noise. Hence, we define the signal-to-noise ratio (SNR) and the signal-to-background ratio (SBR) as follows:

\[ \text{SNR} = 10 \log \frac{P_{\text{signal}}}{P_{\text{GWN}}}, \]  \hspace{1cm} (17)  
\[ \text{SBR} = 10 \log \frac{P_{\text{signal}}}{P_{\text{background}}}, \]  \hspace{1cm} (18)

where \( P \) denotes the power of the subscript argument. The SNR represents the ratio of the strength of the simulated ERPs to that of the added GWN. There is evidence that KF and
Particle filters for ST-ERP estimation 1107

PF are successful in improving the SNR (Doucet et al 2001). The SBR is the ratio of the strength of the desired signal to the added background EEG. Due to the non-Gaussian nature and non-stationarity of the EEGs, the SBR should serve as a realistic criterion. We will discuss the performance of the algorithms later in the presence of each source of noise separately.

The Daubechies-5 wavelet was chosen for its simplicity and general purpose applicability in a variety of time frequency representation problems (Polikara et al 2007). Empirically, we found that three wavelet decompositions should be appropriate for our synthetic and real data sets. The covariance matrix for $w_{k-1}$ and $v_k$ were set to as $Q_w = \sigma_w I$ and $Q_v = \sigma_v I$, in both PF and KF where $I$ is the identity matrix, and $\sigma_w, \sigma_v, q_w, q_v$ are free parameters which are adjusted manually. These parameters determine the performance of the estimator by a trade-off between the estimators’ sensitivity and stability or by a trade-off between noise suppression and signal tracking. In the KF, only the $\sigma_v/\sigma_w$ ratio is important and hence we chose $\sigma_w = 0.1$ for computational convenience. In the PF, $q_v$ and $q_w$ play different roles and their appropriate combination can reduce the estimation error which is defined as the mean square error between the true and estimated ERPs. In this study, we fixed $q_w$ to 5 and only $q_v$ was adjusted. We fixed $q_w$ to simplify the search for the best set of values that optimize our criterion. Note that the performance of the method can increase when both $q_v$ and $q_w$ are tuned rather than when $q_w$ is fixed.

The number of particles in the PF approach was set to 10000. Decreasing the number of particles deteriorates the performance of the PF, and increasing the number of particles causes the computational cost to increase. The computational complexity of the PF is an order of the number of particles, the number of trials and the number of samples in each trial. In contrast, the computational complexity of the KF is an order of the number of trials and the number of samples in each trial. Therefore, the degree to which PF is slower than KF is approximately proportional to the number of particles. Both methods, however, can be applied successfully in real-time processing (online) applications. Figure 1 shows a typical example for the simulated ERPs and their estimations using KF and PF methods. Figure 1(a) depicts the superposition of noiseless ERPs in which the amplitude of the first peak increases while its latency decreases linearly. Figure 1(b) shows the noisy ERPs (simulated EEG) when GWN and background EEG are added to the noiseless ERPs with $\text{SNR} = -5 \text{ dB}$ and $\text{SBR} = -10 \text{ dB}$. Figures 1(c) and (d) show the ERPs extracted from the noisy data using PF and KF methods, respectively. The morphology of ERPs extracted by the PF was closer to the original noise-free ERPs than that extracted by the KF. For instance, note the signal fluctuations in the KF method around 0.25 s. In figure 2, three ST estimations for trial numbers 5, 35 and 55 are shown. Figures 2(a), (c) and (e) demonstrate three different noiseless (thick line) and noisy (solid line) simulated ERPs. The figures show that the ERPs are highly contaminated by the two sources of noise ($\text{SNR} = -5 \text{ dB}$ and $\text{SBR} = -10 \text{ dB}$). Figures 2(b), (d) and (f) show the results of ST estimation of the noisy ERPs using the PF (solid line) and the KF (dashed line) methods. The PF presents more accurate estimates for the peaks and troughs of the noiseless ERP patterns (thick line) than the KF.

Neurophysiologists are primarily interested in peak parameters (amplitude and latency), hence, here these features are carefully modelled in the synthetic ERP data. The results shown in figures 3 and 4 have been obtained by analysis of a set of synthetic data. Other simulations attained comparable or better results. Figure 3 demonstrates true and estimated peak amplitude values on a trial-by-trial basis. Figure 3(a) shows the amplitude of the positive peak for the simulated (thick line) and the estimated amplitudes using the PF and KF methods (solid and dashed lines, respectively). The errors, which are the absolute value of the difference between noise-free and estimated amplitudes in each trial, are plotted in figure 3(c). The mean and variance of errors are plotted in a bar graph in figure 3(d).
Figure 1. Superposition of simulated ERPs and their ST estimations SNR = −5 dB and SBR = −10 dB; (a) simulated ERPs, (b) noisy ERPs (average in thick black line), (c) extracted ERPs using the KF method, (d) extracted ERPs using the PF method. Note that the PF estimates the ERPs very closely with respect to the original ERPs shown in (a).

PF has statistically better performance \((p < 0.001)\). Figures 3(b) and (e) depict the estimated amplitude and the estimated errors for the negative peak. Figure 3(f) shows that the PF has also extracted the second peak amplitudes with statistically smaller error values than the KF \((p < 0.01)\). Figure 4 shows the results of the latency estimation for the positive and negative peaks. The PF reveals the latencies more accurately for both peaks except for the last few trials of the positive peak. Both algorithms, however, give reasonable and indeed similar estimates, and there is no significant difference between their performances.

To further quantify the performance of the methods, they were compared numerically by root mean square error for different levels of SNRs and SBRs. Figure 5(a) shows the output SNR versus input SNR in dB for the Kalman and particle filters. The output SNRs were obtained by search for the best \(\sigma_w\) and \(q_w\) parameters in each SNR in 100 Monte Carlo independent runs. Using the DWT jointly with either PF or KF as a two-staged denoising process results in acceptable performance in high input SNR. Both algorithms, but especially the PF, are robust to GWN and they show consistency for an input SNR as low as \(-10\) dB. Figure 5(b) shows the output SBR versus input SBR for the above methods. The output SBRs were obtained by executing 100 Monte Carlo independent runs (as for the SNR). The PF outperforms the KF for all input SBRs, and both algorithms are more sensitive to background EEG noise than GWN. The output SBR also decreases rapidly in comparison with the output.
SNR. When the amount of background EEG increases and the noise becomes more non-Gaussian, the performance of the PF improves accordingly. Indeed, if the added noise is not Gaussian white noise, which is the case in real EEG data, the PF is more robust than the KF to this violation. It is also noteworthy that, in the PF, any type of noise with known distribution
can be employed. This is in contrast with the KF, where the noise should be GWN with a known covariance matrix.

3.2. Real data results

Read data was obtained in an odd-ball paradigm in the Cognitive Electrophysiology Laboratory, School of Psychology, Cardiff University. Four right-handed healthy individuals participated in the experiment. The experiment was run in a quiet, normally illuminated room. The participants were seated comfortably in an armchair. All gave informed consent.

The electroencephalogram (EEG) was recorded from 25 silver/silver chloride electrodes in an elastic cap and from electrodes on the mastoids. Electrode sites included midline, left and right hemisphere locations (based on the 10/20 system). Eye movements were recorded.
Particle filters for ST-ERP estimation

Figure 4. Latencies of the first and second peaks from trial to trial and their estimations. SNR = −5 dB and SBR = −10 dB; (a) simulated (thick line), estimated using PF (solid line) and estimated latencies using KF (dashed line) for the positive peak, (b) simulated (thick line), estimated using PF (solid line) and estimated latencies using KF (dashed line) results for the negative peak, (c) the error between simulated and estimated latencies for PF (black bar) and KF methods (white bar) for the first peak, (d) mean and first standard deviation of errors for each method, (e) the error between simulated and estimated latencies for PF (black bar) and KF methods (white bar) for the second peak, (f) mean and first standard deviation of errors for each method.

from above and below the right eye (vertical electrooculogram (VEOG)) and the outer canthi (horizontal (HEOG)). The influence of eye blinks was eliminated via a correction procedure based on a linear regression estimate.

During acquisition, the frequency bandwidth of the linear bandpass filter was 0.03–40 Hz, and the sampling rate was 250 Hz. An Fz (midline frontal) reference was employed, and the data were re-referenced (off-line) to the algebraic mean of the signals at the left and right mastoids. A 0.15 s pre-stimulus interval was used for baseline correction.

Each participant heard in total 300 tones, 240 (80%) of which were at one pitch (the frequent tones) and 60 (20%) of which were at a different pitch (the infrequent tones). In this paradigm, ERPs elicited by infrequent tones are associated with a large positive-going wave—the P300—which peaks approximately 0.3 s post-stimulus. Accordingly, epochs from 0.2 s to 0.5 s time-locked to stimulus onset for infrequent trials were extracted for analysis.
By dividing the power of the averaged signal to the mean of power of all trials, an input SBR of approximately $-10\,\text{dB}$ was estimated. By using this SBR and the results obtained from the simulations, the values around $q_w = 5$ and $q_v = 40$ were deemed appropriate. The initial state of the filter was set to zero and, in order to accommodate the convergence of the filter, the first ten trials were excluded. The information contents of those ten first trials can be extracted by running the PF in the reversed trial order. Particle smoothers may also be used as an alternative (Kitagawa 1996). Figure 6 provides the results of the proposed algorithms for real data. The Cz site, at which the P300 component amplitude in the odd-ball paradigm is prominent, was chosen for analysis. Figure 6(a) shows a superposition of all original trials and their average signal (solid line). Original trials are also presented in figure 6(b) in the form of ERP images, in which the ERPs are plotted vertically with time on the horizontal axis. Colour represents the amplitudes of the trials, with blue through red corresponding to the transition from maximum negative through positive amplitudes. The second row of figure 6 shows the results of the PF approach for all of the trials (c), the ERP image (d), and the estimated amplitudes from trial to trial (e). Both the signal and the ERP images show that the P300 has been extracted from noisy ERP data. An increase in the latency and decrease in the amplitude of the P300 over successive trials is evident in figures 6(d) and (e), respectively.
The results of the KF method are shown in figures 6(f)–(h). The KF method does not show structural variations in the amplitude and latency across trials which were identified by the PF.

A primary application of such a single-trial ERP extraction approach can be investigation of the likely induced correlates of the amplitude and latency of ERP components. Figure 7 shows the estimated correlation between ERP amplitudes and latencies using PF (black dots) and KF (red dots) methods for the four subjects. The linear regression has been shown in all figures using black (for the PF) and red (for the KF) lines. There is a significant negative correlation between latency and amplitude of four subjects obtained by the PF ($r = -0.357$ and $p < 0.01$, two tailed) but not by the KF ($r = -0.0986$, $p > 0.05$).
Figure 7. Latency versus amplitude in four subjects with their linear regression obtained by PF shown by black and KF shown by red. A significant negative correlation between amplitude and latency with average $r = -0.357$ ($p < 0.01$, two tailed) for four subjects can be seen only for the PF method ($r = -0.0986$, $p > 0.05$), (a) subject number one (PF: $r = -0.472$, $p < 0.001$) (KF: $r = -0.2864$, $p < 0.05$), (b) subject number two (PF: $r = -0.400$, $p < 0.005$) (KF: $r = 0.0185$, $p > 0.05$), (c) subject number three (PF: $r = -0.227$, $p < 0.05$) (KF: $r = -0.1182$, n.s.), (d) subject number four (PF: $r = -0.327$, $p < 0.05$) (KF: $r = -0.0102$, $p > 0.05$).

4. Conclusions and future directions

We have proposed an approach for ST-ERP estimation based on particle filtering of discrete wavelet transformed ERPs. The main merits of the proposed method are in exploitation of the sequential importance sampling as well as the use of Bayesian theory which promise high performance for non-Gaussian and non-stationary data like ERPs. This method uses only one EEG channel for ST-ERP extraction and, therefore, ignores spatial information contained in EEG data. In addition, as no prior knowledge about the measurement has been considered it may be more sensitive to outliers than other methods such as MAP estimator.

The method was tested for simulated and real data. The simulation results demonstrate the improved accuracy of estimations by PF in comparison with KF, especially when the assumption about GWN is violated (background EEG is considered as noise). Application of the method to real data recorded in an odd-ball paradigm shows that the amplitude of the
P300 decreases and its latency increases over trials during the task. These demonstrations emphasize the potential for this approach in ST analysis of ERP data.

The ability to extract reliably single-trial ERP data would be of great benefit in several contexts. For example, for ERP researchers interested in using ERPs to isolate cognitive processes, the reliance on averaging introduces an inevitable degree of caution when making inferences about the onset times of processes. Caution is also necessary when inferring whether peak amplitude differences between averaged ERPs for different conditions do in fact reflect consistent peak amplitude differences at the level of individual trials. The alternative is that the amplitude differences emerge because of greater inter-trial variability in peak latencies for one condition in comparison to the other.

Similar concerns apply when contrasting ERPs across different patient groups. In this context, the existence of a reliable means of extracting ST-ERPs also offers a means of assessing whether factors such as habituation and fatigue influence ERPs differently according to variables such as disease state or severity, or the location of focal brain damage. Assessment of the performance of the approach described here in populations other than healthy young adults is one way in which this work will be taken forward subsequently.

These observations demonstrate the accuracy of the approach and its potential use in ST-ERP analysis, especially when trial-to-trial variation in ERPs is of major interest. Recently, such ERP estimation has also proved to be very useful in single-trial correlation analysis between EEGs and functional magnetic resonance imaging signals (Bagshaw and Warbrick 2007).

Finally, we are currently developing a robust framework for single-trial dipole source localization based on particle filters. Preliminary results are reported in Mohseni et al (2008a, 2009) where considering the dipole locations as having a nonlinear relation with the measurements, our variational Bayes and maximum likelihood based methods estimate the amplitude, latency and width of the ERPs after they have been localized.

References

Bagshaw A P and Warbrick T 2007 Single trial variability of EEG and fMRI responses to visual stimuli Neuroimage 38 280–92
Cerutti S, Bersani V, Carrara A and Liberati D 1983 Analysis of visual evoked potentials through Wiener filtering applied to a small number of sweeps J. Biomed. Eng. 9 3–12
Liu J S 1996 Metropolized independent sampling with comparison to rejection sampling and importance sampling Stat. Comput. 6 113–9
Mallat S 1999 A Wavelet Tour of signal Processing (New York: Academic)
Mohseni H, Wilding E and Sanei S 2007 Pre-processing of event-related potential signals via Kalman filtering and smoothing 15th IEEE Int. Conf. on Digital Signal Processing (DSP) pp 179–82
Pham D T, Mcks J, Khler W and Gasser T 1987 Variable latencies of noisy signals: estimation and testing in brain potential data Biometrika 74 525–33